Biometrika (1978), **65**, 2, *pp.* 413–8
Printed in Great Britain

A class of parametric models for the analysis of square contingency tables with ordered categories

BY P. McCULLAGH*
Department of Mathematics, Imperial College, London

SUMMARY

A parametric model is developed for the analysis of square contingency tables with ordered categories. Order among the categories is a built-in feature of the new model and this means that it is unnecessary to assign arbitrary 'scores' to the row and column variables. Special cases of the proposed model include conditional symmetry and symmetry. The relationship with marginal homogeneity is also described. The model for quasisymmetry is considered and is shown to be invariant under general permutation transformations of the indices, which makes it suitable for analysing data on a nominal scale. On the other hand, the new model, called p symmetry, is invariant only under the special reverse permutation transformation. This restricted invariance property makes the latter model more suitable for analysing data on an ordinal scale.

Some key words: Conditional symmetry; Marginal homogeneity; Matched pairs; Nominal scale; Ordered categorical data; Palindromic invariance; Quasisymmetry.

1. INTRODUCTION

Square tables of counts frequently arise in medicine, psychology and sociology, where the row variable, X_1, is a measure of mental attitude, health, etc. before treatment, and the column variable, X_2, is the corresponding measurement after treatment. Thus we have a categorized version of the classical matched pairs problem (Koch & Reinfurt, 1971). The objective of such an experiment is to summarize the differences between X_1 and X_2, assumed to be caused by the treatment, taking into account the association between X_1 and X_2.

In general, the categories can be either on a nominal or on an ordinal scale, but the models described in this paper have been developed specifically for ordinal data. The possible conclusions about the treatment effect depend greatly on whether or not the categories are ordered. Thus, if the categories are ordered, a possible conclusion is that X_2 tends to be greater than X_1 or vice versa. Such a statement is meaningless in the context of nominal categories.

The fact that the set of possible conclusions in the case of ordered data is different from the set of possible conclusions for nominal data seems to indicate that models appropriate for one type of data are not generally appropriate for the second type. In the literature for square contingency tables there seems to be little appreciation of the essential difference between ordinal and nominal data. Indeed, the same model is frequently applied to both, although this may be due, in part, to the paucity of models for ordinal data. It is the purpose of this paper to provide a model for ordered categories from which meaningful conclusions, such as those mentioned above, can be drawn.

It is proposed that models for square contingency tables with nominal categories should be invariant under row to column transformations and under arbitrary permutation

* Present address: Department of Statistics, University of Chicago, Illinois.

transformations of the indices. It is understood that the same permutation is applied to both rows and columns. On the other hand, models for ordinal data should not be invariant under general permutations of the indices, but only under the special reverse permutation. That is to say that the model will accept the data in increasing order, $1, \ldots, k$, or in decreasing order, $k, k-1, \ldots, 1$, but will reject all other permutations. Models which satisfy this property are called palindromic invariant.

2. Palindromic invariance

We give here a formal definition of palindromic and permutation invariance although the models can readily be understood and interpreted without reference to their invariance properties. Permutation invariance is formally defined as follows. Let $\Pi(\theta) = \{\pi_{ij}(\theta)\}$ $(1 \leqslant i,j \leqslant k)$ be a model for the k^2 cell probabilities depending on a parameter vector θ in a parameter space Θ. Let $\mathscr{P}$ be the permutation group whose typical element T_α is a $k \times k$ unitary matrix which transforms the vector $J = (1, \ldots, k)^{\mathrm{T}}$ to the permutation $\alpha = (\alpha_1, \ldots, \alpha_k)^{\mathrm{T}}$. The elements $t_{ij}(\alpha)$ of T_α are given by

$$t_{ij}(\alpha) = \begin{cases} 1 & \text{if } j = \alpha_i, \\ 0 & \text{otherwise.} \end{cases}$$

Thus $T_\alpha J = \alpha$ and $T_\alpha^{\mathrm{T}} \alpha = J$.

Let $\mathscr{P}_\nu$ be the subgroup of $\mathscr{P}$ comprising the identity I_k and the reverse permutation operator T_ν which sends J to the reverse permutation $(k, k-1, \ldots, 1)$. The model $\Pi(\theta)$ is said to be permutation invariant if, for every $T \in \mathscr{P}$ and for every $\theta \in \Theta$, there exists a $\theta^* \in \Theta$ depending on θ and T such that

$$T\Pi(\theta)T^{\mathrm{T}} = \Pi(\theta^*). \tag{2·1}$$

Often θ^* is a permutation of the elements of θ but this need not be the case.

The model $\Pi(\theta)$ is said to be palindromic invariant if the only transformations, T, which satisfy (2·1) for every θ are $T \in \mathscr{P}_\nu$; that is $T = I_k$ or $T = T_\nu$.

3. The models

To contrast the different properties of permutation invariance and palindromic invariance we present both models and outline some of their properties. Recall that we are interested in whether or not the row variable, X_1, is stochastically larger than the column variable, X_2. Define

$$\begin{aligned} \pi_{ij} &= \mathrm{pr}(X_1 = i, X_2 = j) \quad (1 \leqslant i,j \leqslant k), \\ P_{ij} &= \mathrm{pr}(X_1 \leqslant i, X_2 \geqslant j) \quad (1 \leqslant i < j \leqslant k), \\ Q_{ij} &= \mathrm{pr}(X_2 \leqslant i, X_1 \geqslant j) \quad (1 \leqslant i < j \leqslant k). \end{aligned}$$

Note that P_{ij} and Q_{ij} have little meaning except in the context of ordered categories. Three models are considered.

Model I: quasisymmetry, q symmetry or QS, where

$$\pi_{ij} = c\frac{\alpha_i}{\alpha_j}\phi_{ij},$$

with $\phi_{ij} = \phi_{ji}$, $\Sigma\Sigma\phi_{ij} = 1$, $\alpha_1 = 1$ and c a constant to make $\Sigma\Sigma\pi_{ij} = 1$. An alternative log linear definition of quasisymmetry is given by Bishop, Fienberg & Holland (1975, p. 268) but the above version is equivalent.

Model II: conditional symmetry, c symmetry or CS, where

$$\pi_{ij} = \begin{cases} \theta\phi_{ij} & (i<j), \\ \phi_{ii} & (i=j), \\ (2-\theta)\phi_{ij} & (i>j), \end{cases}$$

with $\phi_{ij} = \phi_{ji}$ and $\Sigma\Sigma\,\phi_{ij} = 1$.

Model III: palindromic symmetry, p symmetry or PS, where

$$P_{ij} = c\,e^{\frac{1}{2}\Delta}\frac{\alpha_i}{\alpha_{j-1}}\psi_{ij} \quad (1 \leqslant i < j \leqslant k),$$

$$Q_{ij} = c\,e^{-\frac{1}{2}\Delta}\frac{\alpha_{j-1}}{\alpha_i}\psi_{ij} \quad (1 \leqslant i < j \leqslant k), \quad \pi_{ii} = \psi_{ii} \quad (1 \leqslant i \leqslant k),$$

with $\psi_{ij} = \psi_{ji}$, $\alpha_1 = 1$, c chosen to make $\Sigma\Sigma\,\pi_{ij} = 1$, and where $\{\psi_{ij}\}$ satisfy the estimability condition

$$\sum_{i=1}^{k} \psi_{ii} + \sum_{|i-j|=1} \psi_{ij} - \sum_{|i-j|=2} \psi_{ij} = 1, \tag{3·1}$$

which is the analogue of the restriction $\Sigma\Sigma\,\phi_{ij} = 1$ in models I and II.

The invariance properties of the three models are now examined and it is shown that model I is permutation invariant whereas models II and III are palindromic invariant. Thus if, for model I, an arbitrary permutation is applied to both rows and columns, the new cell probabilities π'_{ij} are given by

$$\pi'_{ij} = c\frac{\alpha'_i}{\alpha'_j}\phi'_{ij},$$

where α' is a permutation of the elements of α and ϕ' is obtained from ϕ by permuting both rows and columns. Thus the model for quasisymmetry is permutation invariant.

The reverse permutation applied to model II yields

$$\pi'_{ij} = \begin{cases} (2-\theta)\,\phi_{k-i+1,k-j+1} & (i<j), \\ \phi_{k-i+1,k-i+1} & (i=j), \\ \theta\phi_{k-i+1,k-j+1} & (i>j), \end{cases}$$

and model III becomes

$$P'_{ij} = Q_{k-j+1,k-i+1} \quad (i<j), \qquad Q'_{ij} = P_{k-j+1,k-i+1} \quad (i<j),$$
$$\pi'_{ii} = \pi_{k-i+1,k-j+1} \quad (i=j).$$

It is easy to verify that neither II nor III is invariant under arbitrary permutations and hence they are both palindromic invariant.

The relationships between the models described above and those for symmetry, S, and marginal homogeneity, MH, are now described. In an obvious notation,

(i) S implies CS implies PS implies GPS,
(ii) S implies QS,
(iii) MH with QS is equivalent to S,
(iv) MH with CS is equivalent to S,
(v) MH with PS does not imply S,
(vi) QS with CS is equivalent to S, provided $k > 2$,

where GPS is a generalization of the model for p symmetry and is described below. From the number of constraints imposed by the various models, it also seems that for $k > 3$, QS with

PS is equivalent to S and possibly that, for $k > 5$, QS with GPS is equivalent to S, but these relations are not proved.

We now compare the properties of the three models for making inferences such as that X_1 stochastically is less than X_2 or vice versa. Under quasisymmetry, if $\{\alpha_i\}$ is a nondecreasing sequence, then $\text{pr}(X_1 \leqslant i) \leqslant \text{pr}(X_2 \leqslant i)$ for every i, and hence X_1 is stochastically larger than X_2. However, this implication is not reversible, so that if $\text{pr}(X_1 \leqslant i) \leqslant \text{pr}(X_2 \leqslant i)$ for every i, it does not follow that $\{\alpha_i\}$ is a nondecreasing sequence. Thus if the sequence $\{\alpha_i\}$ is not monotone, X_1 may be stochastically larger than X_2 and hence quasisymmetry is not an appropriate model if we wish to test for stochastic ordering.

Under c symmetry, the corresponding implication is reversible, so that $\theta \leqslant 1$ is equivalent to $\text{pr}(X_1 \leqslant i) \leqslant \text{pr}(X_2 \leqslant i)$ for every i. Similarly, under p symmetry $\Delta \leqslant 0$ is equivalent to $\text{pr}(X_1 \leqslant i) \leqslant \text{pr}(X_2 \leqslant i)$ for every i. Thus, for the inferential viewpoint, the latter models seem preferable to quasisymmetry when the categories are ordered.

Conditional symmetry is a special case of p symmetry obtained by putting $\alpha_i = 1$ $(i = 1, \ldots, k-1)$ and has the conditional interpretation

$$\text{pr}(X_1 = i, X_2 = j \mid X_1 < X_2) = \text{pr}(X_1 = j, X_2 = i \mid X_2 < X_1) = \phi_{ij} \quad (i < j).$$

The model for p symmetry has a similar conditional interpretation,

$$\text{pr}(X_1 \leqslant i,\ X_2 > i \mid X_1 \leqslant i,\ X_2 > i \text{ or } X_2 \leqslant i,\ X_1 > i) = e^{\Delta}/(1 + e^{\Delta}), \tag{3·2}$$

since $P_{i,i+1}/Q_{i,i+1} = e^{\Delta}$. This is also a property of the logistic model for matched ordered categorized data (McCullagh, 1977), where the parameter Δ has a simple quantitative interpretation as the mean difference, on the logistic scale, between X_1 and X_2. The quantity e^{Δ} can also be interpreted as the conditional odds ratio of the event $X_1 \leqslant i$ to the event $X_2 \leqslant i$. The two models are not equivalent but, because of the close similarity, it seems reasonable to interpret Δ as a measure of the lack of marginal homogeneity in both models.

The model for p symmetry has only one parameter measuring lack of marginal homogeneity. Sometimes this will not be sufficient and a more general model is appropriate.

Model IV: generalized palindromic symmetry (GPS), where

$$P_{ij} = c\, e^{\frac{1}{2}\Delta_i} \frac{\alpha_i}{\alpha_{j-1}} \psi_{ij} \quad (1 \leqslant i < j \leqslant k),$$

$$Q_{ij} = c\, e^{-\frac{1}{2}\Delta_i} \frac{\alpha_{j-1}}{\alpha_i} \psi_{ij} \quad (1 \leqslant i < j \leqslant k), \quad \pi_{ii} = \psi_{ii} \quad (1 \leqslant i \leqslant k).$$

The previous restriction (3·1) applies to the parameters $\{\psi_{ij}\}$. The lack of marginal homogeneity is now summarized in the $k-1$ parameters Δ_i $(i = 1, \ldots, k-1)$. All the properties of p symmetry carry over to the more general model with only minor alterations.

Under p symmetry, $\Delta = 0$ implies marginal homogeneity but not symmetry. Similarly, under generalized p symmetry, $\Delta_i = 0$ $(i = 1, \ldots, k-1)$ implies marginal homogeneity but not symmetry. These relationships enable us to test for marginal homogeneity assuming p symmetry or generalized p symmetry. However, the hypotheses of marginal homogeneity and generalized p symmetry are not nested so that MH does not imply GPS. Thus, generalized p symmetry defines a four level hierarchy of models from symmetry through c symmetry to generalized p symmetry.

Each model in the hierarchy has a basic $\frac{1}{2}k(k+1)-1$ parameters for symmetry. The models for c symmetry and p symmetry have an additional 1 and $k-1$ parameters respectively, while that for marginal homogeneity assuming p symmetry has an additional $k-2$

parameters. The generalized p symmetry model has an additional $2k-3$ parameters over the model for symmetry. Thus p symmetry and q symmetry each have $\frac{1}{2}(k+4)(k-1)$ parameters and can be tested for goodness of fit by a chi-squared statistic with $\frac{1}{2}(k-1)(k-2)$ degrees of freedom.

The models for p symmetry and generalized p symmetry are not log linear and there is no quick method of fitting them. Initial estimates of the parameters $\{\Delta_i\}$ can, however, be obtained in an obvious way from the relationship $P_{i,i+1}/Q_{i,i+1} = e^{\Delta_i}$ $(i = 1, \ldots, k-1)$. These are appropriate starting points for the generalized model. In the case of p symmetry these $k-1$ estimates can be combined into a single estimate; see, for example, McCullagh (1977). Furthermore, if $\{\Delta_i\}$ are small and $\{\alpha_i\}$ are close to 1 then $P_{ij}+Q_{ij} \simeq 2c\psi_{ij}$, which can be used to obtain initial estimates of $\{\psi_{ij}\}$. Thereafter, the parameters are estimated by numerical maximization of the log likelihood. When the table is not sparse, this presents no real difficulty and, for the data in the example, convergence was rapid. Problems do arise however when empty cells are symmetrically placed in the table. In this case, the maximum of the log likelihood occurs at a boundary of the parameter space. This may cause numerical difficulties in the maximization procedure. Alternatively, suitably small numbers might be inserted in the empty cells to ensure that the maximum occurs at an interior point of the parameter space.

4. Example

The following example (Stuart, 1953) concerns the unaided distant vision of 7477 women aged 30–39 employed in Royal Ordnance factories in Britain from 1943 to 1946. The row variable, X_1, is the right eye grade and the column variable is the left eye grade. The categories are ordered from highest (1) to lowest (4). The data are presented in Table 1.

Table 1. *Unaided distance vision of* 7477 *women aged* 30–39 *employed in Royal Ordnance factories from* 1943 *to* 1946

		Left eye				
		Highest grade			Lowest grade	
Right eye		(1)	(2)	(3)	(4)	Total
Highest grade	(1)	1520 (0)	266 (1·5)	124 (−9·2)	66 (7·1)	1976
	(2)	234 (−1·6)	1512 (0)	432 (8·6)	78 (−8·8)	2256
	(3)	117 (9·7)	362 (−8·9)	1772 (0)	205 (0·8)	2456
Lowest grade	(4)	36 (−7·7)	82 (9·3)	179 (−0·9)	492 (0)	789
Total		1907	2222	2507	841	7477

Residuals under p symmetry are given in brackets.

The χ^2 value for p symmetry of 6·2 with 3 degrees of freedom indicates a reasonable fit and is a slight improvement over q symmetry: $\chi^2 = 7{\cdot}3$ with 3 degrees of freedom. The estimate of Δ obtained by maximum likelihood is 0·167 with standard deviation 0·046 and indicates that the left eye is significantly worse than the right eye. In addition, model II gives an adequate fit, $\chi^2 = 7{\cdot}3$ with 5 degrees of freedom, with $\hat{\theta} = 1{\cdot}074$, again indicating that the left eye is worse than the right eye. This conclusion contrasts with the corresponding one for 3242 males employed in the same factories. For this data set, the model for symmetry fits adequately: $\chi^2 = 4{\cdot}76$ with 6 degrees of freedom. The estimated value of Δ for the 3242 males is $-0{\cdot}064 \pm 0{\cdot}068$. Thus there is no evidence of marginal inhomogeneity in the distance vision data for the males.

To demonstrate the lack of invariance of p symmetry under arbitrary permutations of the indices, the analysis of Table 1 was repeated with rows 1 and 2 and columns 1 and 2 interchanged. The resulting χ^2 value of 12·9 with 3 degrees of freedom indicates a poor fit, as would be expected.

Finally, we note, without explanation, the strange residual pattern under p symmetry of the data in Table 1. The same residual pattern occurs when the model for quasisymmetry is fitted; see, for example, the fitted values given by Plackett (1974, p. 61). However, the same residual pattern does not occur with the data for the 3242 male employees.

I am grateful to Professor D. R. Cox and to Dr A. C. Atkinson for their helpful comments and to Mr A. T. Jenkinson who suggested the name palindromic symmetry. This research was supported by the Health and Safety Executive.

References

Bishop, Y., Fienberg, S. E. & Holland, P. W. (1975). *Discrete Multivariate Analysis: Theory & Practice*. Massachusetts Institute of Technology Press.

Koch, G. G. & Reinfurt, D. W. (1971). The analysis of categorical data from mixed models. *Biometrics* **27**, 157–73.

McCullagh, P. (1977). A logistic model for paired comparisons with ordered categorical data. *Biometrika* **64**, 449–53.

Plackett, R. L. (1974). *The Analysis of Categorical Data*. London: Griffin.

Stuart, A. (1953). The estimation and comparison of strength of association in contingency tables. *Biometrika* **40**, 105–10.

[*Received April* 1977. *Revised November* 1977]